

## Entrega 7. Proyecto FST - Hasta el 27 de mayo de 2022 a las 23:59

Escribir el código Python, utilizando la librería Pynini, para transformar un texto en HTML al un texto en Markdown teniendo en cuenta las siguientes restricciones:

- El texto HTML a transformar se leerá del fichero de entrada llamado "texto.html". El texto transformado en Markdown se almacenará en un fichero con el mismo nombre que el fichero de entrega, pero con la extensión ".md" (MiApellido\_MiNombre.md).
- Solo se considerarán los elementos de HTML que se indican a continuación. Los elementos que no se consideren se eliminarán en el texto transformado. Los atributos de las etiquetas, si los hubiera, no se tendrán en consideración.
- Suponga que el texto está formado únicamente por caracteres alfanuméricos. En principio, la ñ y las letras acentuadas no las consideraremos ya que se codifican (o se pueden codificar) con una entidad propia de HTML (&acute; &ntilde;). No obstante, puede añadir al alfabeto los signos de puntuación típicos, por ejemplo: .,:;()?!". Así como el espacio en blanco, el tabulador (\t) y la nueva línea (\n).
- Suponga que el texto HTML está bien formado.
- **[A]** El elemento párrafo (<p>...texto...</p>) se transformará en un párrafo Markdown, es decir, se escribirá el mismo contenido finalizándolo con un doble retorno de carro (...texto...\n\n).
- **[A]** El elemento de salto de línea (<br>) se transformará en un salto de línea Markdown, es decir, la cadena formada por dos espacios seguida del retorno de carro (.. \n).
- **[A]** El elemento que representa una línea horizontal (<hr>) se transformará en una línea (o regla) horizontal Markdown, es decir, la cadena formada por tres guiones (---).
- **[A]** Los elementos de encabezado (<h1>...titulo1...</h1>, <h2>...titulo2...</h2>, ..., <h6>...titulo6...</h6>) se transformarán en los correspondientes encabezados Markdown, es decir, usando almohadillas, una por nivel, precediendo la texto del título (# ...titulo1..., ## ...titulo2..., ..., ##### ...titulo6...). Puede restringir la transformación de los elementos de encabezado a h1, h2 y h3, no considerando el resto.
- **[A]** Los elementos de énfasis, cursivas (<i>...texto...</i>) y negritas (<b>...texto...</b>), se transformarán en la notación Markdown correspondiente: para la cursiva encerrando el texto entre \* (...texto...\*) y para la negrita encerrando el texto entre \*\* (\*\*...texto...\*\*). Además, hay que tener en cuenta que si queremos usar los dos tipos de énfasis podemos combinar la sintaxis para poner el texto en cursiva y negrita (\*\*...texto...\*\*).
- **[B]** Los elementos de listas desordenadas (<ul> <li>...e1...</li> <li>...e2...</li> ... </ul>) se transformarán en una lista desordenada Markdown (- ...e1..., - ...e2..., ...). No se considerarán ni las lista anidadas ni las listas ordenadas.
- **[C]** El elemento de enlace en línea (<a href= "...enlace...">...texto\_enlace...</a>) se transformarán en enlaces en notación Markdown: escribiendo el texto enlazado entre corchetes y el link en cuestión entre paréntesis ([...texto\_enlace...](...enlace...)).

Por ejemplo, el texto de entrada:

```

<html>
<h2>Titulo</h2>
<h3>Subtitulo</h3>
<p>Al texto en Markdown podemos incluirle formato como
<b>negrita</b> o <i>cursiva</i> o <i><b>ambas</b></i> de una manera
muy sencilla.</p>
<p>Este es un ejemplo de texto que da entrada a una <b><i>lista
generica</i></b> de elementos:
<ul>
<li>Elemento 1</li><li>Elemento 2</li>
<li>Elemento 3</li></ul>
</p>
<hr>
<p>Enlace a la pagina principal de la <a
href="http://www.uva.es">Universidad de Valladolid</a>.</p>
</html>

```

Se debe transformar en:

```

## Titulo
### Subtitulo
Al texto en Markdown podemos incluirle formato como negrita o
cursiva o ambas de una manera muy sencilla.
Este es un ejemplo de texto que da entrada a una lista
generica de elementos:
- Elemento 1
- Elemento 2
- Elemento 3
---
Enlace a la pagina principal de la [Universidad de
Valladolid] (http://www.uva.es) .

```

La puntuación del proyecto vendrá determinada por la puntuación de cada una de las partes, concretamente:

- **[A]** 6 puntos.
- **[B]** 2 puntos.
- **[C]** 2 puntos.

### Entrega.

Escriba el código Python que permita la transformación del texto bien en un único archivo de texto (extensión txt), bien en un único archivo ipynb (cuaderno Python/cuaderno Jupyter), comprímalo con ZIP y renómbrelo a MiApellido\_MiNombre.zip, usando su primer apellido y su nombre (si alguno de los dos es compuesto use el primero de los términos, si comienza por 'de' o 'de la', elimine estas palabras). Se entregará este archivo comprimido.